Evaluating ADR Programs

A Handbook' for

Federal Agencies

Administrative Conference of the United States

Dispute System Design Working Group

March 1995

HANDBOOK ON ADR PROGRAM EVALUATION

INTRODUCTION

What is ADR program evaluation? Simply stated, it is a way to determine whether an alternative dispute resolution (ADR) program is meeting its goals and objectives. Evaluation provides data and analysis to answer questions such as:

- Is the ADR program, as implemented, consistent with its original design?
- Is the program reaching the intended audience(s)?
- How well did the program achieve its goals and objectives? (e.g. What is the impact of the program on dispute resolution cost, time, and outcomes?)
- Are participants satisfied with program administration and outcomes?

Evaluation data are useful in finding out what works and what does not in an ADR program, and may be a critical factor in decisions to modify or expand a program.

This handbook highlights issues that arise in designing and carrying out an evaluation of an ADR program. The issues have been set out in a question and answer format. Some areas of evaluation design, addressed only generally here, are highly technical in nature, and require specialized expertise. These include development of vaed and reliable data collection instruments, sampling methodologies, and certain kinds of data analysis, particularly statistical methodologies. Alternative dispute resolution program staff and others involved in ADR evaluation efforts who do not have technical expertise in these and related matters should consult appropriate experts both before and throughout the evaluation process.

A few important points to keep in mind:

- ADR program evaluation should be responsive to the needs of those who have a stake in the program, e.g. users, managers, decisionmakers, and others, and should relate to the ADR program's goals.
- Evaluation is an "art," as well as a science. Evaluators need to strike a workable balance between the need for defensible results and practical limitations.
- Evaluation is also not a "linear' process. Decisions made early in the evaluation planning and design process may need to be reconsidered and modified appropriately.
- Flexibility is important throughout the evaluation process.

Systematic planning for program evaluation at the time an ADR program is set up allows appropriate data collection mechanisms to be established early on, thereby avoiding the potentially difficult and time-consuming process of trying to reconstruct useful data.

It is hoped that this handbook will help ADR practitioners and program administrators focus on issues involved in ADR program evaluation and ensure that evaluation results will be useful.

PLANNING THE EVALUATION

The planning stage of an evaluation involves identifying and refining the evaluation's goals and objectives. Useful planning steps include:

- determining goals and objectives for the evaluation;
- identif .ying the audience(s) for the evaluation, and how best to meet their needs;
- considering issues of timing and expense; and
- selecting an evaluator.

What are the goals and objectives of the evaluation?

There are a variety of remns to evaluate an ADR program. Program managers may want to know how weU the program is working or whether-chmges should be made. There may be questions about whether the program should be continued or expanded. There may be an interest in finding out how well ADR works in particular federal contexts. Or, a program evaluation may be required by Congress or agency management.

The goals and objectives of the evaluation should reflect the needs and interests of those requesting the evaluation, and should be sensitive to the needs and interests of the expected audience(s) for the results. Tne evaluation's goals should be closely tied to the goals and objectives of the program being evaluated. Ideally, the ADR program's goals and objectives wiU have been established in its design phase. Sometimes, however, these goals may not have been clearly articulated, may not be measurable as stated, or may

have changed. Evaluators may need to ask program managers and other stakeholders to provide input (and hopefuuy arrive at a consensus) on the program's goals.

ne Perfonnance Indicators for ADR Program Evaluation (November 1993), prepared by the Adnu'm'strative Conference's Dispute Systems Design Worldng Group, and attached as Appendix A, describes some of the more common goals of ADR programs, as well as indicators of success (what to measure to determine whether program goals have been met). The document may be helpful to those seeking to define or select ADR program goals.

Who are the audiences for the evaluation and what do they want to know?

There are usuafly a variety of people who have an interest in the results of a program evaluation. These audiences may be interested in different issues and seek different types of information. Potential audiences should be identified as early as possible, and kept in mind in planning the evaluation, so that their questions will be addressed. Possible audiences for an ADR program evaluation include ADR program officials, other agency officials, program users, Members of Congress, and others. Program officials may be interested in finding out how the ADR program is working, and how it might be improved. Their interests might focus, for example, on the program's impact on case inventory (backlogs), the effects of ADR use on long-term relationships among disputants, or how well information about the program is being disseminated. Program officials involved in the day-to-day operation may have different interests than those at higher levels.

Other agency officials**,** including budget officers, staff within Offices of General Counsel and Inspector General, or managers from other programs, may also have an interest in evaluation results. Budget officials may be interested in whether cost savings have been achieved through implementation of the program. The Inspector General may be interested in the nature of the settlements and whether ADR use promotes long-term compliance. General Counsels may care about how long it takes to resolve cases or the nature of outcomes; other managers may want to know how effectively the program was implemented.

Members of Congress and their staffs may be interested in how ADR use affects budgets and how related laws, such as the Administrative Dispute Resolution Act, are being implemented. Members of the public may be interested in how efficiently the agency is resolving its disputes, and how satisfied participants are with ADR processes. Disputants may be interested in finding out how typical their experience was compared to other users. Officials In other federal agencies may find evaluation results helpful as they plan or modify-their own ADR programs. There may well be other audiences whose interests or desire for information should be considered.

While it is not possible to answer all potential questions, it is useful to figure out what the possible questions are and then focus the evaluation on the most important ones. Talking to members of the various potential audiences can help identify the issues they are

interested in, and may help develop consensus about which issues to address. Such discussions also improve the likelihood that evaluation results will be a useful and meaningful part of future decisionmaking processes.

How will evaluation results be used?

Although terminology differs, evaluations are commonly characterized as either. (1) program effectiveness (also known as impact, outcome or summative) evaluations, which focus on whether a program is meeting its goals and/or having the desired impact; or (2) program design and administration (also known as process or formative) evaluations, which examine how a program is operating. Program effectiveness evaluations often may be useful in determining whether a program should be continued or expanded; program design/administration evaluations often focus on how a continuing program can be improved.

Remember that decisions on the future of programs (or even how they could be improved) are usually not made solely on the basis of program evaluation results. Agency priorities, other institutional concerns, budget limitations, and other factors will also affect program decisions.

When should a program evaluation be undertaken?

Evaluation can be undertaken at different times during the life of an ADR program. Among the factors to consider are (1) whether the program has been in operation long enough to ensure that there are sufficient cases to look at; (2) whether the program has gotten the early bugs out; (3) if this is a pilot, whether the evaluation will be completed early enough to be a factor in the decision to continue or expand the program; and (4) whether there are other deadlines relating to future decisionmaidng which will affect the usefulness of evaluation results.

Timing decisions in any particular evaluation depend on the evaluation's goals and objectives. An evaluation aimed at measuring the program's effectiveness should wait until there is a realistic potential for a measurable impact. While a program design evaluation (aimed at improving a program) could in principle be done at any time, a program effectiveness evaluation may be most useful if start-up bugs have been eliminated.

Who will conduct the evaluation?

In selecting an evaluator (or team of evaluators), a number of qualifications should be considered. Objectivity (i.e., no stake in the outcome) is a critical quaURcation. In addition, experience in conducting program evaluations is very helpful. Sufficient technical expertise is very important in designing the data collection process and analyzing the datl Such expertise may be found, for example, inside some agency policy and program evaluation offices, at the U.S. General Accounting Office, or at various outside evaluation consulting firms and university departments specializing in social

science research. Some understanding of the organization or the context in which the program operates can be helpful to the evaluator, as are good interpersonal and management skills.

Evaluations can be conducted by people outside the agency, within the agency but outside the program being evaluated, or by people involved with the ADR program. There are advantages and disadvantages to each option. An outside evaluator has the potential for the greatest impartiality, which can be important for valid and credible results. In addition, depending on the expertise available in a particular agency, an outside evaluator may have more technical knowledge and experience. Outside evaluation may be relatively expensive, however, depending on the affiliation of the evaluators (e.g. colleges or universities, other non-profit groups, or private sector entities such as management consulting or social science research f=s). If the agency has evaluation capacity outside the organization where the ADR program is being implemented, the requisite neutrality may be available at a potentially lower cost. Using someone involved in the program implementation or design may be the least expensive, and offer the best understanding of program context, but it also carries with it potential perceptions of a lack of impartiality. One way to avoid some of the disadvantages of each of these approaches is to use a team of people, representing internal and external groups.

It may be useful to set up a group of stakeholders as an informal "advisory committee," which can be used as a sounding board on evaluation design and implementation issues, to help refine the focus of the evaluation, and to review and comment on data collection methodologies as the evaluation proceeds.

Regardless of who does the evaluation (outside or inside), it is useful to have someone in the ADR program who can serve as a liaison with the evaluator to ensure access to the necessary information. The liaison might be the person responsible for planning the evaluation.

How much will the evaluation cost?

The cost of conducting an ADR program evaluation depends on a number of factors, including:

- number and complexity of performance indicators selected;
- type of design selected;
- level of statistical significance required of results;
- availability of acceptable data for comparison purposes; and
- who is selected to carry out the evaluation.

Costs can be controlled, however, through careful planning and appropriate adjustments in the design phase.

What are some other constraints on evaluation and how can they be overcome?

The evaluation team needs to identify constraints on evaluation and develop strategies for dealing with them. In addition to budget and resource constraints, organizational opposition or concerns and operational difficulties may also need to be, overcome. A successful evaluation requires the cooperation and support of policymakers, ADR program managers and staff, and those from whom data will be sought. Any anticipated lack of cooperation needs to be identified and addressed. Depending on the design of the evaluation, issues relating to user access to the program may need to be considered. For example, an evaluation based on a controlled experiment will deny access to some potential users of the program, which may not be feasible in a particular context.

How long will it take to conduct the evaluation?

There is no single answer to this question. It may take weeks or months, depending on the complexity of the planned evaluation, the size of the program and sample, whether data have been gathered on an ongoing basis, and whether the evaluation team has other significant demands on its time. There must be time for planning and designing the evaluation, for collecting the data, for analyzing it, and for preparing a report. Evaluators should try to establish target dates and timelines to the extent possible.

DESIGN AND IMPLEMENTATION

Evaluation design is a subject widely covered in the program evaluation literature. Because this handbook is aimed specifically at an ADR audience, however, many of whom are not familiar with the technical aspects of research design, we have laid out a set of broad steps for designing and implementing an effective ADR program evaluation. They are as follows:

- Gain a thorough understanding of ADR program design and operation.
- Translate evaluation goals and objectives into measurable performance indicators.
- Determine data needs and availability of data.
- Determine appropriate design strategy.
- Decide how to collect the data.
- Collect the data.
- Analyze and interpret the data.

The material that follows further elaborates on each of these elements.

What does the ADR program currently look like?

Evaluators need to gain a thorough understanding of the ADR program's design and implementation (initial design and actual operation). Such understanding is important for selecting appropriate performance indicators, determining an appropriate design strategy, assessing data availability, and selecting data collection methods. Knowledge of the program also allows evaluators to ask appropriate questions and work effectively with others involved in the evaluation process. Further, complete and accurate description allows those using the evaluation results to understand the context.

Reviewing documents and interviewing key program managers and stakeholders is a good way to get an understanding of the ADR program. The types of documents evaluators review at this stage may include applicable laws and regulations, program budgets and expenditure reports, written program plans and policies, and organization charts. The following kinds of questions should be considered in this analysis:

- What is the program supposed to accomplish; i.e., what are its goals and objectives? How are goals and objectives defined and prioritized? How much agreement is Lhere on program goals and objectives?
- Can program goals be measured? Have measurable indicators been established? Are data already being collected to measure those objectives?
- Who are the program participants (disputants and other participants)?
- How are cases referred to the program?
- What functions and responsibilities are given to program participants? How are their activities coordinated?
- What limits are placed on the authority of the program or program participants? Are there other constraints to program operation?
- What funding and staffing resources are available and being used for the program?
- What type of training is provided to program staff and participants? Are there written policies and procedures for program operation?
- What types of program data are avglable? Can the data be used to evaluate program results? If not, what additional information would be nee@@, and what is the estimated time and cost of gathering it?

What performance measures/indicators should be selected?

Once goals and objectives for the evaluation have been established (during the planning process) and a thorough understanding of program content and administration gained, evaluators are in a position to establish appropriate measures or performance indicators. Performance indicators represent the questions being asked in the evaluation, and serve as the basis for data collection and analysis. They should be well-defined, as clear and specific as possible, and collectively address the issues of concern to program managers, participants, other stakeholders, and decisionmakers.

As noted earlier, the Performance Indicators for ADR Program Evaluation (attached as Appendix A) provides a large number of sample indicators. A given ADR program will probably have multiple goals, and many indicators will be of potential interest. It is not easy or necessary to measure all aspects of an ADR program; some types of data are harder to obtain and some areas of a program are more important to examine than others. Moreover, absent unlimited time and resources, it is usually better to focus on answering a few key questions. Once performance indicators have been selected, evaluators will be ready to select a design strategy, identify data needs, and determine data collection strategies.

What design strategy should be used?

A variety of research designs are available for use in evaluating ADR programs. Research designs are simply strategies for determining who or what will be the subject of the research and when. A particular design is selected based on the type of evaluation being conducted (e.g. effectiveness vs. process evaluations), the type of data needed, data availability, the timeftwne for evaluation, budgetary constraints, and other factors.

Possible design strategies for ADR program evaluation include case studies, time series designs, and comparison group designs. Case studies focus on single, cohesive groups or cases, e.g., a group of cases in which ADR was elected, and are more descriptive in nature. Any comparisons that might be made, for example, to similar situations in which ADR might not have been elected, are inferential and hold no scientific weight This is because the case study design does not allow causes other than the experimental treatment-in this case, ADR-to be ruled out.

Time series designs involve the collection of information about a particular group over periods of time. These designs enable evaluators to examine performance or behavior over time. Again, there is no separate comparison group.

Comparison group designs divide individual persons or, cases at the beginning of the study into different groups, those participating in the program (the experimental group), and those who are not (the control group). The control group serves as a comparison for evaluating the behavior of the experimental group. Comparison groups are useful for determining whether outcomes are the result of using ADR or have some other cause, because any other change in the environment would affect both groups.

"True experimental' and 'quasi-experimental' designs both involve the use of comparison groups; however, in the case of true experimental designs, individuals or cases are assigned randomly to experimental and control groups. Random assignment is the best way to ensure that outcomes can indeed be attributed to the program being implemented. This is particularly important if an evaluation is aimed at measuring the effectiveness of an ADR program. It is not always possible or even desirable to assign individuals or cases randomly, however. For example, it may be inappropriate to Emit access to an ADR pr6jim. Evaluators can then mm to the 'next best thing'-quasi-experimental designs. Quasi-experimental designs also use comparison groups, but assignment to the groups is not random. Rather, evaluators use naturally occurring groups that are as similar as possible, for example, a group of cases disposed of before the ADR program began is compared to a group of cases handled under the ADR program.

There are technical names for design variations within these basic categories of design methodology. Evaluation planners should consult persons with research methodology expertise when selecting a design strategy. Different strategies are appropriate in different situations, and evaluation needs and constraints vary. A rigorous design may be especially appropriate when the evaluation is conducted in-house rather than by outside evaluators. However, not every evaluation needs to be of the rigorous, 'tr-ue experimental' sort in order to have value to the evaluation consumer; nor is such rigor always possible. Evaluators may not have the time, resources, or ability to control access

to the program being studied. In all cases, however, choosing the best possible design strategy will enhance the quality and credibility of evaluation results.

What data do you need?

Once you know what questions you want to answer, what you want to measure, and who (or what) will be measured at what time, the next consideration is what data to couect. In answering this question, you need to consider in advance how the data will be analyzed (see discussion below), to make sure that the appropriate data are collected.

The purpose of this section is to suggest types of data that can be useful in evaluating ADR programs. In general, there are certain categories of data that need to be considered:

Program Data: Evaluators must be able to describe both how the program is intended to operate and how it operates in practice. Information about the program's design may be available from legislation, regulations, other agency documents, and program officials and designers. Information describing the program's actual operation may be obtained through direct observation, interviews, or program records. For example, an ADR program may have written criteria specifying the types of cases eligible for the ADR program. Through interviewing program staff or reviewing program records, evaluators - may -find - out - whether additional or different criteria were being used for selecting cases. Interviews can also be conducted with program users, e.g. disputants and disputant representatives, to determine how a program is actually operating.

Comparison Data: In a program effectiveness evaluation, evaluators need to be able to describe what the situation would have been without the program (sometimes referred to as the baseline), to serve as a basis for comparison. For example, to determine whether agency costs are reduced using ADR, evaluators might examine all or a sample of cases to determine how much it costs to handle disputes without ADR. This information could then be compared to data reflecting how much it costs to handle disputes *using* ADR. Comparison information is not always easy to obtain, especially if the agency does not keep ongoing records reflecting the relevant qw of information. For example, not all agencies keep information about the costs of administering a particular type of program, or the amount of time agency personnel spend on particular types of tasks that may contribute to an ADR program.

Ideally, comparison data are collected by using a control group; that is, a group that looks like the group that went through the ADR program but which did not use ADR. Since a control group is not always available, alternatives must sometimes be;6sed, although they may not provide as good a comparison. For example, comparison data can be collected by looking at the historical period before the ADR process was started. (See earlier discussion of time series and other designs.)

Information relating to possible alternative causes: It is possible that results or outcomes of a program are the result not of the program itself, but of some other factor or event that has taken place during the program implementation. A few possible alternative causes or

explanations for particular results from an ADR program might include changes in agency policy or regulations governing the subject of the dispute or affecting settlement policy, changes in the economic environment that may affect the outcomes of disputes, or the particular slciu of the individuals participating in the implementation of a pilot program. It is important to look for and be able to rule out alternative causes.

How do you collect data?

Once the evaluators have identified the questions the program evaluation is intended to answer and the data needed to answer those questions, the next consideration is how the data should be collected. There are several different sources of data, and several different methods of collecting them. The most common sources are documents, individuals, and observations.

Data may be collected from agency or program documents. These might include plans, procedures, case files, accounting records, budget documents, contracts, computerized data bases, and existing studies, reports, and secondary analyses of the ADR program. It is important to identify the types of documents that need to be reviewed and the data to be obtained from them. In addition, consider whether all records, or just a sample, will be reviewed. If a sample is to be used, it should be representative of the program caseload. (NOTE: Where sampling of persons or records is to be used, the evaluation team must use appropriate sampling procedures. Consult the experts when members of the evaluation team do not have the-requisite technical knowledge or expertise.)

Evaluators can develop checklists, coding sheets, - or standardized forms to help them simplify and systematize document review. These help ensure that evaluators review records in a similar way and collect similar data, and can enhance the reliability of the data collected.

Data can be collected from people. Potential sources include agency and program staff, program participants, subject matter experts, and other stakeholders with an interest in the program. Such testimonial data are generally collected through individual interviews, questionnaires or *focus* groups. There are several methods for obtaining information from people.,

Interviews

> *Standardized interviews* are conducted using an interview guide. Interview guides outline the issues and topics to be explored and help ensure that comparable data are gathered from everyone interviewed. They can be as informal as a "tickler" fist of questions, or as formal as a structured questionnaire. Interview guides can incorporate open-ended questions that allow the interviewer to do some probing or they can use fixed, close ended questions that require the respondent to choose from among several fixed answers.

Informal conversational interviews tend to be unstructured and spontaneous. They may be part of an ongoing participant observation exercia such as watching a mediation.

Telephone interviews are conducted by administering the interview guide or questionnaire to the respondent over the telephone. This technique can be used to conduct structured interviews, with evaluators asking the same questions of numerous individuals in a precise manner, or to conduct more informal interviews.

Focus group interviews allow evaluators to collect information from a group of people simultaneously and allow for interaction and exchange of ideas among the participants. Evaluators usually use a moderator to guide and focus the discussion

Written surveys

Participant surveys are questionnaires administered to program clients or participants immediately or shortly after program services are provided.

Mail moseys are questionnaires that are administered through the mail. Response rates are often low, and evaluators cannot be certain that the person to whom the survey was directed actually filled it out. However, they tend to be relatively inexpensive.

Data can also be collected through observation or inspection of activities, properties, or events. An example might be observation of a mediation session. Evaluators can use rating schemes or less structured ways of observing and documenting the process. Be aware that outside observers may affect the interaction being observed.

At the behest of the Administrative Conference, RAND prepared a set of prototype data collection materials for use by federal agencies in ADR program evaluations. These and other examples of surveys and questionnaires are useful starting points for designing such materials. It is important to make appropriate modifications to meet the needs of the program being evaluated.

How do you select people to survey?

There are several methods for selecting individuals to interview or survey. Evaluators might select interviewees based on their position or role in designing, implementing, or evaluating the program. They might also look to key informants who are particularly knowledgeable about the program or a subject matter.

In some cases, all affected people can be questioned. In other cases, only a sample will be used. Formal types of sampling include random sampling, where evaluators select individuals at random based on a listing, or "sampling frame," and purposeful sampling, where evaluators select individuals, on the basis of some characteristic, as being

representative of a particular view or subgroup. Evaluators might also ask others to name participants or stakeholders in various program activities. As suggested earlier, consultation with a statistician or research methodologist will help, evaluators select the method most appropriate to meeting the objectives of the evaluation.

The key point to remember is that if not all participants are being surveyed, the ones that are must be representative of the group at large. Otherwise, the data may be skewed or biased.

How can you make sure that you will be collecting good quality -data?

There are a variety of factors that may have an impact on the quality of the gathered data. Data must be measured in ways that are reliable and valid. A measure is considered reliable if repeated applications of the measurement instrument such as a survey question, yield essentially the same results. For example, a file review checklist or coding sheet may be considered reliable if two evaluators use it to review the same set of files and obtain similar results. Validity focuses on the accuracy of measurement. A measure is considered valid if it produces a true representation of the event being measured. For example, reviewing ADR case records may be a valid way of measuring the costs of ADR to participants, but only if the records accurately reflect all the participant costs associated with using ADR to handle disputes. Incomplete or inaccurate data have limited usefulness for drawing evaluation conclusions. There are a number of threats to the validity of data that relate to how and when measurements are made. Among these are:

> *Instrumentation* effect-when the observed effect is due to changes made in a recording instrument such as a questionnaire or checklist, between one use of the instrument and another, or where the data collection instrument itself affects the responses (e.g., by using leading questions).
>
> *Hawthorne* effect-when the observed effect is due to factors associated with the experiment itself, such as awareness of the importance of a study, the newness of a prograta or staffing the program with higher-than-usual-quality staff.
>
> *Selection bias-when* the observed effect is due to differences between the types of individuals in the sample and in the comparison group.

There are other, more technical issues that can arise, depending on the scientific rigor expected of an evaluation. Research methodologists can help evaluators design the evaluation to minimize these threats to validity.

What are some things to be aware of in collecting data?

There are a number of issues that need to be taken into account in designing and implementing data collection. Some of these are discussed here.

Confidentiality of data: Often agency records contain names of individuals who are entitled to privacy and should not be named in the evaluation report. Similarly, when collecting information through interviews, evaluators may want to keep confidential the names of people supplying testimonial information. Offering confidentiality to interviewees may increase their willingness to cooperate with the evaluation. Be aware that circumstances may exist that require breach of confidentiality. The law does not provide a special privilege for surveys, nor are they automatically exempt from disclosure under the Freedom of Information-Act. Although anonymous surveys should protect the identity of respondents, absolute confidentiality should not be promised unless it can delivered.

Paperwork Reduction Act.- Evaluators of federal agency ADR programs need to consider the requirements of the Paperwork Reduction Act when designing data collection instruments (see 5 CFR § 1320.7). The Act provides that agencies "shall not conduct or sponsor the collection of information" without first obtaining approval from the Office of Management and Budget. "Collection of information" includes soliciting facts or opinions through the use of written report forms and questionnaires from ten or more persons (from .within or outside the federal government). It may cover oral interviews based on a written interview guide.

Uniformity.- It is important that data be collected in a uniform manner. For example, if more than one person is conducting interviews, they all need to be asking the same questions, in the same order. Procedures need to be developed and training provided to ensure that information is being collected in a reliable manner that permits comparison of the data.

Pretesting. It is *essential* to pretest data collection instruments, to ensure that questions are clear and can be understood by those who will answer them, and that they provide the information the evaluator needs to do the planned analysis.

How are data analyzed?

Strategies for analyzing data are an essential component of an evaluation. Data analysis involves describing and explaining data, and may include comparing groups, identifying patterns and trends, and establishing relationships among variables. Consultation with statisticians or survey research experts may be necessary to ensure that appropriate data collection instruments, survey samples, and analytical techniques are used. There are some initial considerations that should be addressed before selecting appropriate analytical techniques.

- Will the audience appreciate and understand sophisticated analyticd techniques?
- How much margin of error in numerical estimates will the audience feel comfortable with?
- Is generalization from a sample to a population desirable?
- Are -evaluators -attempting to show a -causal relationship between the program and measured results?

The evaluation's objectives and audience expecta6ons will be integral to selecting an appropriate data analysis strategy. Such strategies include simple descriptive methods and more complex statistical methods. The amount of time available to complete the report may rule out extensive data collection and analysis options. Consultation with research methodologists will help in determining the best data analysis strategy for a particular study. Automated statistical packages are available to ease the process of recording and analyzing data. They are especially useful when a research design involves extensive surveying of program participants. Just remember, data analysis is a critical part of the evaluation process, since the findings, conclusions, and recommendations, if any, must be based on the analysis.

PRESENTATIONG DISSEMINATION, AND USE OF RESULTS

Presenting and disseminating the results of your ADR program evaluation is an important stage of the process. Results should be communicated in ways that will allow meaningful decision making by program administrators and decisionmakers.

It is easier to make decisions about the best way to present and disseminate results if the people who will be using the results (the audience) have been consulted during the evaluation process. Such consultation can avoid costly or embarrassing errors; e.g., omission of a key area for analysis, and can ensure the report meets the needs of those who will be using it.

What is the best method for communicating your results?

There are a variety of ways that evaluators can communicate results to potential audience(s). Evaluators may provide briefings, hold meetings with users, and/or prepare a written report.

> *Briefings and presentations* allow evaluators to convey important evaluation information quickly and selectively. In selecting material to be presented, care should be taken to avoid bias or presentation of material out of context. Some discussion of methodology is important as are appropriate cautions about the limits and appropriate use of evaluation data. Providing for interaction with or feedback from the audience may allow issues and potential problems to be identified.
>
> *Written reports* typically take a great deal of time to prepare, but allow evaluators to provide considerably more detail on both methodology and results. A final written report is often required by legislation or executive decision. If it is important to ensure that there is one "official" source of information on evaluation methodology and results, a formal written report can serve this purpose well.

What kind of information needs to be communicated?

Although the potential audiences, program content, and evaluation objectives will vary for each ADR program evaluation, it is generally helpful to include the following kinds of information in a report or other type of presentation:

- Goals and objectives of the evaluation
- Description of the ADR program and how it operates
- Description of the evaluator's methodology
- Presentation of evaluation findings
- Discussion of program strengths and weaknesses
- Implications for program administration (e.g. training, etc.)
- Recommendations as appropriate

Presentation style is entirely a matter of what works for whom. It is always important, however, to make sure that evaluation data are presented accurately and completely, to prevent charges of misrepresentation or overreaching, and to avoid misuse of results.

How can you enhance the effectiveness of your presentation?

Variations in presentation format and style aside, we offer the following suggestions for making the presentation of evaluation results as effective as possible.

Involve potential users as early as possible in determining presentation form and style. Evaluation data should be organized and communicated in a way that is useful for potential audiences and users.

Tailor presentation method, format and style to audience needs. Select the method of presentation (e.g. oral briefing, written report), format, and style of presentation (e.g. formal vs. informal, briefing vs. discussion) based on who your audience is and what their needs are. There may be multiple audiences with multiple needs. Be flexible and willing to adapt material as appropriate.

Be clear and accurate. Evaluation information must be presented clearly and accurately. Always keep the audience in mind as you prepare to describe the ADR program and present evaluation data. Avoid any gaps in describing the program or presenting the results. A clear and accurate portrayal of the ADR program and ADR evaluation results will allow the audience to draw appropriate conclusions about program effectiveness and any need for change.

Be honest and direct. Sharing evaluation findings with potential users and involving them in key decisions concerning presentation format and style does not mean publishing only those findings that reflect well on the program or those affiliated with it. Evaluators must present the story objectively; too heavy an emphasis on the positive may cast doubt on the integrity of the results as well as the integrity of the evaluator(s). Data that suggest weaknesses in program design or administration or that reveal failure to accomplish program goals or objectives should be reported and can be used as a basis for suggesting appropriate changes.

Honest analysis and thoughtful consideration of the information will enhance both the credibility and usefulness of the results.

Keep the body of the report or the bulk of the presentation simple,- reduce complex data to understandable form; use graphic illustrations where appropriate. Evaluation results must be presented so that the most essential data are available, understandable, and useful. Too complex a format or over reliance on narrative may detract from evaluation results and analysis. Organize the presentation or report for multiple uses. Use headings and subheadings to help the audience identify useful information quickly. Limit the use of technical jargon. Prevent misinterpretation or misuse by considering what the data will look like if lifted from the context of the presentation or report. Use simple graphics to illustrate results and call attention to key findings. Make footnotes and technical data

available in handouts or appendices so that the body of the presentation or report is as uncomplicated as possible.

Provide an executive summary or abstract. Evaluators should provide an overview. The 'quick take' should be supplemented by more detailed discussion later in the report.

Make survey instruments and other data collection tools available. Materials can be made available as handouts, at an oral presentation or face-to-face meeting, or as appendices to a written report. The availability of such material enhances both understanding and credibility. It also allows other ADR program evaluators to learn from the experiences of their peers.

Note limitations on the interpretation and use of evaluation data, where appropriate Limitations on the interpretation of the data, such as those that might relate to the replicability of study results, should be communicated to the audience. Evaluators need to exercise caution in expressing their own views and conclusions. Where conclusions are not an objective reflection of the data, they need to be labeled appropriately; i.e., as the views of the evaluator(s) and not necessarily of officials responsible for the program.

Expect the need for follow-up: be flexible and responsive. Have extra copies of reports and presentation handouts available. Keep materials accessible. Provide addresses and telephone numbers for follow-up discussion or questions. Be available for consultation. Stay abreast of how results are being used; provide clarification or added direction in the case of misinterpretation or misuse. Prepare additional materials as needed. Tailor subsequent releases to customer needs.

Who is responsible for making decisions regarding the dissemination of evaluation results?

It is important to think about dissemination of the results at two points: early in the planning process, and again as results become available. Decisions about dissemination may be made solely by the evaluator, solely by program officials or other entity that has requested the evaluation, or, more typically, cooperatively. Such decisions may be circumscribed by contract or agreement, or may be discussed and resolved informally by evaluators and decisionmakers.

When should evaluation results be made available?

Decisionmakers need to consider the implications of releasing evaluation results at different times. For example, if you want publicity for the results, select slower news days. You may want to coordinate release with other agency activities. The timing of data release may be defined by contract or agreement, or may otherwise be discussed and resolved by evaluators and decisionmakers. Releasing preliminary data before all data are collected or analyzed may be risky.

How widely will evaluation results be disseminated?

Evaluation results may be disseminated widely or narrowly. Cost, convenience, and level of interest are likely to play a role. It is rare that either the evaluator or program officials will have complete control over dissemination of the results.

How will evaluation results be initially disclosed?

Evaluation results can be initially disclosed in different ways, with more or less fanfare. They may be made available to the selected audience(s) by memorandum, by press release, by press conference, etc. Typically, such decisions will be made at the executive level, by those who have the authority to make the disclosure.

Performance Indicators for ADR Program Evaluation

Background Information

This section is intended to serve as guidance on the identification of both program goals and program measures. Program goals and measures are really two sides of the same coin. Goals are what your program seeks to accomplish; measures are used to determine whether those goals have been met. The material contained in this section can be used directly to identify possible measures of success for ADR programs. It can therefore be used at both the "front" and "back" ends of program planning and implementation (design and evaluation, respectively).

Evaluations are conducted for different reasons and take different forms. Evaluation may be aimed at (1) determining whether the outcomes of a program are consistent with the program's declared goals, (2) determining whether the program is running the way it was intended to, and/or (3) determining whether changes in the program would improve its usefulness. Evaluations may be comprehensive in nature, rely on a significant degree of internal or external professional evaluation expertise, involve a great deal of planning, and take a rather lengthy time to complete. At the other end of the spectrum, evaluations may be aimed at providing more of a "snapshot" of where a program is, at examining a particular area within a program, or at capturing the impact of specific changes in program coverage or administration. They may involve less planning and outside evaluation expertise, and take a relatively short period of time to complete. Or, the nature and form of an evaluation may fall somewhere in between these two ends of the spectrum. The reasons for which an evaluation is conducted, and the form it takes, will vary from agency to agency, and from time to time, depending on evaluation needs and constraints (e.g., budgetary), and each agency's particular mission/culture. Evaluations need to be designed to be responsive to managers and decisionmakers with different needs and interests.

The list of indicators below is divided into two categories, dealing with program effectiveness (i.e., whether or not a program is meeting its goals), and one dealing with program design and administration (i.e., whether or not a program is being administered as it should be). (Note: the terms "program measures" and "performance indicators" are used interchangeably throughout this section to refer to specific ways of examining program effectiveness or administration.) These categories are not mutually exclusive, and the list itself is intended to be as comprehensive as possible, in order to cover a wide range of agency interests/needs. It is unlikely that all of the measures listed below would apply to any single evaluation; rather, some will apply in some cases and others, not. Each measure is followed by one or more questions intended to further illustrate the kinds of evaluation issues an agency may wish to pursue.

The list, overall, is intended simply as a "sampling" of measures or indicators from which agencies may pick and choose, as appropriate, as they seek to formulate ADR program goals and to identify possible measures of program effectiveness.

List of Indicators

Program Effectiveness

- Efficiency

  Cost

  Time

- Effectiveness
    - Dispute Outcomes
    - Durability of Dispute Outcomes
    - Impact on Dispute Environment
- Customer Satisfaction
    - Participants' Satisfaction with Process
    - Impact on Relationships Between Parties
    - Participants' Satisfaction with Outcomes

Program Design and Administration

- Program Organization
    - Program Structure and Process
    - Directives, Guides and Standards
    - Delineation of Responsibilities
    - Sufficiency of Staff
    - Coordination/Working Relationships
- Service Delivery
    - Access and Procedure
    - Case Selection Criteria
- Program Quality
    - Training
    - Neutrals
- Other Specific Program Features

A. Program Effectiveness (Impact)

Program effectiveness measures or indicators are aimed at assessing the degree to which an ADR program is meeting its goals. More specifically, program effectiveness measures are used to examine the impact of the program on users/participants, overall mission accomplishment, etc. In the case of ADR, an agency may, for example, be interested in looking at whether the use of ADR reduces the time it takes to resolve cases/disputes.

Effectiveness indicators should correspond directly to the goals or objectives of an ADR program. For example, if a goal of your agency's ADR program is to reduce the backlog of cases, than the impact of the program on case disposition time needs to be assessed.

The indicators in the effectiveness category are further divided into three subcategories: efficiency, effectiveness, and customer satisfaction.

Efficiency

1. Cost

- Cost to the Government of using ADR vs. traditional dispute resolution processes (e.g. negotiated settlements, agency findings, litigation). *Is the use of ADR more or less costly than the use of traditional means of dispute resolution? (Cost may be measured in staff time, dollars, or other quantifiable factors.)*

- Cost to disputants of using ADR vs. traditional dispute resolution processes. *Is the use of ADR more or less costly than the use of traditional means of dispute resolution? (Cost may be measured in terms of staff time, dollars, or other quantifiable factors.)*

2. Time

- Time required to resolve disputes using ADR vs. traditional means of dispute resolution. *Are disputes resolved more or less quickly using ADR processes, compared to traditional means of dispute resolution? Such factors as administrative case processing, participant preparation, dispute resolution activity time frames, and/or days to resolution may be considered.*

Effectiveness

1. Dispute Outcomes

- Number of settlements achieved through the use of ADR vs. traditional dispute resolution processes. *Does the use of ADR result in a greater/fewer number of settlements?*

- Number of cases going beyond ADR steps. *Does the use of ADR result in a greater/fewer number of investigations, further litigation activities, etc. ?*

- Nature of outcomes. *What impact does the use of ADR have on the nature of outcomes, e.g. do settlement agreements "look different," as in terms of the agreement or monetary amounts agreed upon? Do settlement agreements reflect more "creative" solutions? Do outcomes vary according to the type of ADR process used*

- Relationship, for cases selected for ADR, between dispute outcomes and such factors as complexity or number of issues, or number of parties. Is *there any relationship, where ADR is used, between the complexity and/or number of parties/issues in a case and the outcome of the case?*

2. Durability of Outcomes

- Rate of compliance with settlement agreements. *Does the use o ADR result in greater/lesser levels of compliance with settlement agreements?*

- Rate of dispute recurrence. *Does the use of ADR result in greater/lesser levels of dispute recurrence, i e. recurrence of disputes among the same parties?*

- Impact on program/organizational environment. *Does use of ADR have the effect of improving the work environment, e.g. reducing the level of' conflict and improving participant relationships, thereby contributing positively to mission accomplishment?*

3. Impact on Dispute Environment

- Size of case inventory. *Does the use of ADR result in an increase/decrease in case inventory?*

- Types of disputes. *Does the use of ADR have an impact on the types of disputes that arise?*

- Negative impacts. *Does the use of ADR have any negative consequences, e.g. an inability to diagnose and correct systemic problem/issues?*

- Timing of dispute resolution. *Does the use of ADR affect the stage at which disputes are resolved?*

- Level at which disputes are resolved. *Does the use o fADR have any impact on where and by whom disputes are resolved?*

- Management perceptions. W*hat are the quantitative and qualitative qffects of using ADR on management, e.g. how does the use of ADR impact upon allocation*

and use of management time and resources? Does the use of ADR ease the job of managing?

- Public perceptions. *Is the public satisfied with ADR outcomes? Is there any perceived impact of use of ADR on effectiveness of the underlying program? (NOTE: "Public" may be defined differently, depending on the particular program/setting involved)*

Customer Satisfaction

1. Participants' Satisfaction with Process

- Participants' perceptions of fairness. W*hat are participant perceptions of 'access to ADR, procedural fairness, fair treatment of parties by neutrals, etc. ?*

- Participants' perceptions of appropriateness. *What are participant perceptions of appropriateness of matching decisions (i.e. matching of particular ADR processes to particular kinds of disputes or specific cases) ?*

- Participants' perceptions of usefulness. *What are participant perceptions of the usefulness of ADR in the generation of settlement options, the quantity and reliability of information exchanged, etc. ?*

- Participants' perceptions of control over their own decisions ("destiny.") *Do participants feel a greater or lesser degree of control over dispute resolution process and outcome through the use of ADR? Is greater control desirable?*

2. Impact on Relationships Between Parties

- Nature of relationships among the parties. *Does the use of ADR improve or otherwise change the parties' perceptions of one another? Is there a decrease/increase in the level of conflict between the parties? Are the parties more or less likely to devise ways of dealing with future disputes? Are the parties able to communicate more directly/effectively at the conclusion of the ADR process and/or when new problems arise?*

3. Participant's Satisfaction with Outcomes

- Participants' satisfaction with outcomes. *Are participants satisfied/unsatisfied with the outcomes of cases in which ADR has been* u*sed?*

- Participants' willingness to use ADR in the future. *Would participants elect to use ADR in a future dispute(s)?*

B. Program Design and Administration (Structure and Process)

How a program is implemented will have an impact on how effective a program is in meeting its overall goals. Program design and administration measures or indicators are used to examine this relationship and to determine how a program can be improved.

The indicators in the program design and administration category are further subdivided into three subcategories: program organization, service delivery, and program quality.

Program Organization

1. Program structure and process.

- *Is the program structure and process consistent with underlying laws, regulations, executive orders, and/or program guidance? Do program structure and process adequately reflect program design? Is the program structure and process adequate to permit appropriate access to and use of the program?*

1. Directives, guides, and standards.

- *Do program directives, guides, and standards provide staff/users with sufficient information to appropriately administer/use the program?*

3. Delineation of responsibilities.

- *Does the delineation of staff/user responsibilities reflect program design? Is the delineation of responsibilities such that it fosters smooth and effective program operation?*

4. Sufficiency of staff (number/type).

- *Is the number/type of program staff consistent with program design and operational needs?*

5. Coordination/working relationships.

- *Is needed coordination with other relevant internal and external individuals and organizations taking place? Have effective working relationships been established to carry out program objectives?*

Service Delivery

1. Access and procedure

- Participant access to ADR option. *Are potential participants made aware of the ADR program? Is the program made available to those interested in using ADR?*

- Relationship between participant perceptions of access and usage of ADR. *What impact does participants perceptions about the availability of the program have on the levels of program usage?*

- Participant understanding of procedural requirements. *Do program users understand how the program works? Did they feel in advance that they were comfortable with the process?*

- Relationship between procedural understanding and rates of usage. *Is there any relationship between the level of participant understanding and the degree of program use, e.g. is a lack of participant understanding serving as a disincentive to using the ADR program?*

2. Case Selection Criteria.

  - Participant perceptions of fairness, appropriateness. *Do participants feel that appropriate types of cases are being handled in the ADR program? Do participants or non-participants feel that the criteria for which cases are eligiblefor ADR are fair? Are cases being sent to the ADR program at the appropriate dispute stages?*

  - Relationship between dispute outcomes and categories of cases. *Is there a correlation between the (nature, size, types of disputants, and/or stage of the dispute) of cases and the outcome of the dispute? Are certain types of cases more likely to be resolved through ADR than others?*

Program Quality

1. Training.

- Participant perceptions of the appropriateness of staff and user training. *Do participants feel that they were provided with sufficient initial information and/or training on how to use the ADR program process? Do they feel that program staff had sufficient training and/or knowledge to appropriately conduct the ADR program?*

- Relationship between training variable and dispute outcomes. *Is there a relationship between the type/amount of training (for participant and/or staff and dispute outcomes?*

2. Neutrals

- Participant views of the selection process. *Are participants satisfied with the manner in which neutrals were selected and assigned to cases? Where they involved in the selection decision? If not, did they feel they should be?*

- Relationship between participant views of the selection process, perceptions of neutral competence and objectivity, and dispute outcomes. *Is there any relationship between participant views about the selection process and dispute outcomes? How do these views affect participants' assessment of the competence and neutrality of neutrals?*

- Participants perceptions of competence (including appropriateness of skill levels/training). *Do participants feel that neutrals were sufficiently competent? Do participants feel that neutrals were sufficiently well trained? Do participants feel that more or less training was needed?*

- Participant perceptions of neutrality/objectivity. *Do participants feel that neutrals were sufficiently objective? Do participants feel that neutrals were fair in their handling of the dispute?*

Other Specific Program Features

Every ADR program is unique. Those requesting and/or conducting an evaluation may want to consider examining other aspects of the ADR program. These unique features may relate to the design of a program, who was and continues to be involved in program design and administration, etc. Each is likely to have at least some impact on service delivery and the quality of the program and should be considered for inclusion in either a comprehensive or selected evaluation of the program, as appropriate.